

A Simplified Approach to Recovery Conditions for Low Rank Matrices

Samet Oymak[†], Karthik Mohan[‡], Maryam Fazel[‡] and Babak Hassibi[†]

[†] California Institute of Technology
{soymak, hassibi}@caltech.edu

[‡] University of Washington, Seattle
{karna, mfazel}@uw.edu

Abstract—Recovering sparse vectors and low-rank matrices from noisy linear measurements has been the focus of much recent research. Various reconstruction algorithms have been studied, including ℓ_1 and nuclear norm minimization as well as ℓ_p minimization with $p < 1$. These algorithms are known to succeed if certain conditions on the measurement map are satisfied. Proofs for the recovery of matrices have so far been much more involved than in the vector case.

In this paper, we show how several classes of recovery conditions can be extended from vectors to matrices in a simple and transparent way, leading to the best known restricted isometry and nullspace conditions for matrix recovery. Our results rely on the ability to “vectorize” matrices through the use of a key singular value inequality.

Index Terms—rank minimization, sparse recovery

I. INTRODUCTION

Recovering sparse vectors and low-rank matrices from noisy linear measurements, with applications in compressed sensing and machine learning, has been the focus of much recent research. The Restricted Isometry Property (RIP) was introduced by Candès and Tao in [5], [3] and has played a major role in proving recoverability of sparse signals from compressed measurements. The first recovery algorithm that was analyzed using RIP was ℓ_1 minimization in [5], [3]. Since then, many algorithms including GraDes [16], Reweighted ℓ_1 [24], and CoSaMP [25] have been analyzed using RIP. Analogous to the vector case, RIP has also been used in the analysis of algorithms for *low rank matrix recovery*, for example Nuclear Norm Minimization [27], SVP [21], Reweighted Trace Minimization [23] and AdMiRA [20]. Other recovery conditions have also been proposed for recovery of both sparse vectors and low-rank matrices including the Null Space Property [30], [26] and the Spherical Section Property [30], [12] (also known as the ‘almost Euclidean’ property) for the nullspace. The first matrix RIP result was given in [27] where it was shown that the RIP is sufficient for low rank recovery using nuclear norm minimization, and that it holds with high probability as long as number of measurements are sufficiently large. This

analysis was improved in [4] to require a minimal order of measurements. Recently, [23] improved the RIP constants with a stronger analysis similar to [1].

In this paper, we show that if a set of conditions are sufficient for the robust recovery of sparse vectors with sparsity at most k , then the “extension” (defined later) of the same set of conditions are sufficient for the robust recovery of low rank matrices up to rank k .

While the recovery analysis in [23] and [4] (Theorem 2.4) is complicated and lengthy, our results (see “Main Theorem”) are easily derived due to the use of a key singular value inequality (Lemma 1). Our results also apply to generic recovery conditions, one of which is RIP. As an example, $\delta_k < 0.307$ [1], $\delta_{2k} < 0.472$ [2] are two of the many RIP-based conditions (see also [3], [15]) that are known to be sufficient for sparse vector recovery using ℓ_1 minimization. A simple consequence of this paper is the following: The RIP conditions $\delta_k < 0.307, \delta_{2k} < 0.472$ (and all other RIP conditions that are sufficient for sparse vector recovery) are *also* sufficient for robust recovery of matrices with rank at most k improving the previous best condition of $\delta_{2k} < 0.307$ in [23]. Improving the RIP conditions is a direct consequence of our observation but is not the focus of this paper, although such improvements have been of independent interest (e.g., [1], [2]).

Our results also apply to another recovery condition known as the Nullspace Spherical Section Property (SSP) and it easily follows from our main theorem that the spherical section constant $\Delta > 4k$ is sufficient for the recovery of matrices up to rank k as in the vector case [30]. This approach not only simplifies the analysis in [12], but also gives a better condition (as compared to $\Delta > 6k$ in [12]). Our final contribution is to give nullspace based conditions for recovery of low-rank matrices using Schatten- p quasi-norm minimization, which is analogous to ℓ_p minimization with $0 < p < 1$ for vectors. These nonconvex surrogate functions have motivated algorithms such as IRLS [22], [10] that are empirically observed to improve on the recovery performance of ℓ_1 and nuclear norm minimization.

This work is [†]supported in part by the National Science Foundation under CCF-0729203, CNS-0932428 and CCF-1018927 and [‡]supported in part by NSF CAREER grant ECCS-0847077.

II. BASIC DEFINITIONS AND NOTATION

For a vector $\mathbf{x} \in \mathbb{R}^n$, $\|\cdot\|_0$ denotes its cardinality or the number of nonzero elements. $\bar{\mathbf{x}}$ denotes the vector obtained by decreasingly sorting the absolute values of the entries of \mathbf{x} , and \mathbf{x}^k denotes the vector obtained by restricting \mathbf{x} to its k largest elements (in absolute value). Let $\text{diag}(\cdot) : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^n$ return the vector of diagonal entries of a matrix, and $\text{diag}(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ return a diagonal matrix with the entries of the input vector on the diagonal. Let $n_1 \leq n_2$. We denote rank of a matrix $X \in \mathbb{R}^{n_1 \times n_2}$ by $\text{rank}(X)$, and its i th largest singular value by $\sigma_i(X)$. Let $\Sigma(X) = [\sigma_1(X), \dots, \sigma_{n_1}(X)]^T$ be vector of decreasingly sorted singular values of X . X^k denotes the matrix obtained by taking the first k terms in the singular value decomposition of X . The nuclear norm of X is denoted by $\|X\|_* = \sum_{i=1}^{n_1} \sigma_i(X)$, and its Frobenius norm by $\|X\|_F = \sqrt{\sum_{i=1}^{n_1} \sigma_i^2(X)}$. Let $\Sigma_X = \text{diag}(\Sigma(X)) \in \mathbb{R}^{n_1 \times n_1}$. We call (U, V) a unitary pair if $U^T U = U U^T = V^T V = I$. In this paper, we'll use the following singular value decomposition of X : $X = U \Sigma_X V^T$ where (U, V) is a unitary pair. Clearly $U \in \mathbb{R}^{n_1 \times n_1}$, $V \in \mathbb{R}^{n_2 \times n_1}$. Notice that the set of matrices $U D V^T$, where D is diagonal, form an n_1 dimensional subspace. Denote this space by $S(U, V)$. Let $\mathcal{A}(\cdot) : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ be a linear operator. $\mathcal{A}_{U,V}(\cdot) : \mathbb{R}^{n_1} \rightarrow \mathbb{R}^m$ is called the restriction of \mathcal{A} to unitary pair (U, V) if we have $\mathcal{A}_{U,V}(\mathbf{x}) = \mathcal{A}(U \text{diag}(\mathbf{x}) V^T)$ for all $\mathbf{x} \in \mathbb{R}^{n_1}$. In particular, $\mathcal{A}_{U,V}(\cdot)$ can be represented by a matrix $A_{U,V}$.

Consider the problem of recovering the desired vector $\mathbf{x}_0 \in \mathbb{R}^n$ with $\|\mathbf{x}_0\|_1 = k$ from corrupted measurements $\mathbf{y} = A\mathbf{x}_0 + \mathbf{z}$, with $\|\mathbf{z}\|_2 \leq \epsilon$ where ϵ denotes the noise level, and $A \in \mathbb{R}^{m \times n}$ denotes the measurement matrix. Under certain conditions, \mathbf{x}_0 can be found under certain conditions by solving the following convex problem,

$$\begin{aligned} & \text{minimize} && \|\mathbf{x}\|_1 \\ & \text{subject to} && \|A\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \end{aligned} \quad (1)$$

where recovery is known to be robust to noise as well as imperfect sparsity. We say \mathbf{x}^* is *as good as* \mathbf{x}_0 w.r.t \mathbf{y} if $\|A\mathbf{x}^* - \mathbf{y}\|_2 \leq \epsilon$ and $\|\mathbf{x}^*\|_1 \leq \|\mathbf{x}_0\|_1$. In particular, the optimal solution of the problem 1 is as good as \mathbf{x}_0 .

Similarly, consider the case where the desired signal to be recovered is a low-rank matrix denoted by $X_0 \in \mathbb{R}^{n_1 \times n_2}$, with $n = n_1 \leq n_2$. Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ be the measurement operator. We observe corrupted measurements $\mathbf{y} = \mathcal{A}(X_0) + \mathbf{z}$ with $\|\mathbf{z}\|_2 \leq \epsilon$.

$$\begin{aligned} & \text{minimize} && \|X\|_* \\ & \text{subject to} && \|\mathcal{A}(X) - \mathbf{y}\|_2 \leq \epsilon. \end{aligned} \quad (2)$$

Similar to the vector case, we say that X^* is *as good as* X_0 w.r.t \mathbf{y} if $\|\mathcal{A}(X^*) - \mathbf{y}\|_2 \leq \epsilon$ and $\|X^*\|_* \leq \|X_0\|_*$. In particular the optimal solution to *problem 2* is as good as X_0 . We now give the definitions for certain recovery conditions on the measurement map, the Restricted Isometry Property and the Spherical Section Property.

Definition: Restricted Isometry Constant (RIC). *The RIC*

of a matrix A is the smallest constant δ_k for which

$$(1 - \delta_k) \|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1 + \delta_k) \|\mathbf{x}\|_2^2 \quad (3)$$

holds for all vectors \mathbf{x} with $\|\mathbf{x}\|_0 \leq k$.

The RIC for linear operators, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is defined similarly, with X instead of \mathbf{x} , \mathcal{A} instead of A and $\text{rank}(X) \leq k$ instead of $\|\mathbf{x}\|_0 \leq k$, and $\|X\|_F$ instead of $\|\mathbf{x}\|_2$.

Definition: Restricted Orthogonality Constant (ROC). *The ROC of a matrix A is the smallest constant $\theta_{k,k'}$ for which*

$$|\langle A\mathbf{x}, A\mathbf{x}' \rangle| \leq \theta_{k,k'} \|\mathbf{x}\|_2 \|\mathbf{x}'\|_2 \quad (4)$$

holds for all vectors \mathbf{x}, \mathbf{x}' with disjoint supports and $\|\mathbf{x}\|_0 \leq k$ and $\|\mathbf{x}'\|_0 \leq k'$.

Our definition of ROC for linear operators, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is slightly looser than the one given in [23]. For a linear operator \mathcal{A} , ROC is the smallest constant $\theta_{k,k'}$ for which

$$|\langle \mathcal{A}(X), \mathcal{A}(X') \rangle| \leq \theta_{k,k'} \|X\|_F \|X'\|_F \quad (5)$$

holds for all matrices X, X' such that $\text{rank}(X) \leq k$, $\text{rank}(X') \leq k'$ and both column and row spaces of X, X' are orthogonal, i.e., in a suitable basis we can write $X = \begin{bmatrix} X_{11} & 0 \\ 0 & 0 \end{bmatrix}$ and $X' = \begin{bmatrix} 0 & 0 \\ 0 & X'_{22} \end{bmatrix}$.

We say that the matrix A (or operator \mathcal{A}) satisfies **RIP** if the corresponding RIC and ROC constants are sufficiently small to ensure recovery of sparse vectors (low rank matrices). As one example if a matrix A satisfies $\delta_k < 0.307$ or $\delta_{2k} < 0.472$, then we say A satisfies RIP, since these conditions are sufficient for recovery of k -sparse vectors using ℓ_1 minimization.

Definition: Spherical Section Property (SSP). *The Spherical Section constant of a linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ is defined as*

$$\Delta(\mathcal{A}) = \min_{Z \in \mathcal{N}(\mathcal{A}) \setminus \{0\}} \frac{\|Z\|_*^2}{\|Z\|_F^2},$$

and we say \mathcal{A} satisfies the Δ -Spherical Section Property if $\Delta(\mathcal{A}) \geq \Delta$.

The definition of SSP for a matrix $A \in \mathbb{R}^{n \times m}$ for analyzing recovery of sparse vectors is analogous to the above definition. Another way to describe this property is to note that a large Δ implies the nullspace is an *almost Euclidean* subspace [30], whose elements cannot have a small ratio of ℓ_1 (nuclear norm) to ℓ_2 (Frobenius norm), and therefore the subspace cannot include sparse (low-rank) elements.

In the subsequent sections, we give our main results relating vector recovery to matrix recovery.

III. KEY OBSERVATIONS

Throughout this note, many of the proofs involving matrices apply the following useful Lemma which enables us to “vectorize” matrices when dealing with matrix norm inequalities.

Lemma 1. ([18]) (**Key Lemma**) For any $X, Y \in \mathbb{R}^{n_1 \times n_2}$, we have

$$\sum_{i=1}^{n_1} |\sigma_i(X) - \sigma_i(Y)| \leq \|X - Y\|_*. \quad (6)$$

We now give an application of Lemma 1.

Lemma 2. Given W with singular value decomposition $U\Sigma_W V^T$, if there is an X_0 for which $\|X_0 + W\|_* \leq \|X_0\|_*$ then there exists $X_1 \in S(U, V)$ with $\Sigma_{X_1} = \Sigma_{X_0}$ such that $\|X_1 + W\|_* \leq \|X_1\|_*$. In particular this is true for $X_1 = -U\Sigma_{X_0}V^T$.

Proof: From Lemma 1 we have

$$\|X_0 + W\|_* \geq \sum_i |\sigma_i(X_0) - \sigma_i(W)|. \quad (7)$$

On the other hand, for $X_1 = -U\Sigma_{X_0}V^T$, W we have

$$\|X_1 + W\|_* = \|-\Sigma_{X_0} + \Sigma_W\|_* = \sum_i |\sigma_i(X_0) - \sigma_i(W)|. \quad (8)$$

Then from (7) and (8) it follows

$$\|X_1 + W\|_* \leq \|X_0 + W\|_* \leq \|X_0\|_* = \|X_1\|_*. \quad (9)$$

■

Although Lemma 2 is easy to show, it proves helpful in Theorem 1 to connect vector recovery (on restricted subspaces $S(U, V)$) to matrix recovery over all space $\mathbb{R}^{n_1 \times n_2}$.

To further illustrate the similarity between sparse and low rank recovery, we state the null space conditions for noiseless recovery.

Lemma 3. ([9], [30]) **Null space condition for sparse recovery**

Let $A \in \mathbb{R}^{m \times n}$ be a measurement matrix. Assume $\epsilon = 0$ then one can perfectly recover all vectors \mathbf{x}_0 with $\|\mathbf{x}_0\|_0 \leq k$ via program 1 if and only if for any $\mathbf{w} \in \mathcal{N}(A)$ we have

$$\sum_{i=1}^k \bar{w}_i < \sum_{i=k+1}^n \bar{w}_i, \quad (10)$$

where \bar{w}_i is the i th entry of $\bar{\mathbf{w}}$ defined previously in section II.

The use of the above condition has a long history; it was stated in [11] (see also [13], [14]) for matrices made from concatenation of two bases, and studied in [9], [30] in a general setting.

Lemma 4. ([26]) **Null space condition for low-rank recovery** Let $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ be a linear measurement operator. Assume $\epsilon = 0$ then one can recover all matrices X_0 with $\text{rank}(X_0) \leq k$ via program 2 if and only if for any $W \in \mathcal{N}(\mathcal{A})$ we have

$$\sum_{i=1}^k \sigma_i(W) < \sum_{i=k+1}^{n_1} \sigma_i(W). \quad (11)$$

IV. MAIN RESULT

In this section, we state our main result which enables us to seamlessly translate results for vector recovery to matrix recovery. Our main theorem assumes that operator \mathcal{A} satisfies an extension property, defined below.

Definition: Extension. Let \mathbf{P} be a property defined for matrices, $A \in \mathbb{R}^{m \times n}$. We say that a linear operator, $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ satisfies the extension property, \mathbf{P}^e if all its restrictions $A_{U,V}$ have property \mathbf{P} .

RIP and SSP are two examples of property \mathbf{P} that are of interest in this paper. We show later that the main theorem can be applied to these properties.

Let $\|\cdot\|_v$ be an arbitrary norm on \mathbb{R}^n with $\|\mathbf{x}\|_v = \|\bar{\mathbf{x}}\|_v$ for all \mathbf{x} . Let $\|\cdot\|_m$ be the corresponding unitarily invariant matrix norm on $\mathbb{R}^{n_1 \times n_2}$ such that $\|X\|_m = \|\Sigma(X)\|_v$. For the sake of clarity, we use the following shorthand notation in the main theorem for statements regarding recovery of vectors:

- V_1 : A matrix $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ satisfies a property \mathbf{P} .
- V_2 : In program 1, for any \mathbf{x}_0 , $\|\mathbf{z}\|_2 \leq \epsilon$, $\mathbf{y} = A\mathbf{x}_0 + \mathbf{z}$ and any \mathbf{x}^* as good as \mathbf{x}_0 we have,

$$\|\mathbf{x}^* - \mathbf{x}_0\|_v \leq h(\bar{\mathbf{x}}_0, \epsilon). \quad (12)$$

for some h .

- V_3 : For any $\mathbf{w} \in \mathcal{N}(A)$, \mathbf{w} satisfies a property \mathbf{Q} .

We also use the following shorthand for statements regarding recovery of matrices:

- M_1 : A linear operator $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ satisfies the extension property \mathbf{P}^e .
- M_2 : In program 2, for any X_0 , $\|\mathbf{z}\|_2 \leq \epsilon$, $\mathbf{y} = \mathcal{A}(X_0) + \mathbf{z}$ and any X^* as good as X_0 we have,

$$\|X^* - X_0\|_m \leq h(\Sigma(X_0), \epsilon). \quad (13)$$

- M_3 : For any $W \in \mathcal{N}(\mathcal{A})$, $\Sigma(W)$ satisfies property \mathbf{Q} .

Theorem 1. (Main Theorem)

For a given \mathbf{P} , the following implications hold true:

$$(V_1 \implies V_2) \implies (M_1 \implies M_2) \quad (14)$$

$$(V_1 \implies V_3) \implies (M_1 \implies M_3) \quad (15)$$

Proof: $(V_1 \implies V_2) \implies (M_1 \implies M_2)$.

Assume $V_1 \implies V_2$ and that M_1 holds. Consider program 2 where measurements are $\mathbf{y}_0 = \mathcal{A}(X_0) + \mathbf{z}_0$ with $\|\mathbf{z}_0\|_2 \leq \epsilon$. We would like to show that for any X^* that is as good as X_0 w.r.t \mathbf{y}_0 , it holds that $\|X^* - X_0\|_m \leq h(\Sigma(X_0), \epsilon)$. Consider any such X^* and let $W = X^* - X_0$. This implies that $\|X_0 + W\|_* \leq \|X_0\|_*$ and $\|\mathcal{A}(X_0 + W) - \mathbf{y}_0\|_2 \leq \epsilon$. Then, from Lemma 2 for $X_1 = -U\Sigma_{X_0}V^T$ (where W has SVD $U\Sigma_W V^T$) we have $\|X_1 + W\|_* \leq \|X_1\|_*$. Now, let $\mathbf{y}_1 = \mathcal{A}(X_1) + \mathbf{z}_0$. Clearly

$$\|\mathcal{A}(X_1 + W) - \mathbf{y}_1\|_2 = \|\mathcal{A}(X_0 + W) - \mathbf{y}_0\|_2 \leq \epsilon \quad (16)$$

and hence $X_1 + W$ is as good as X_1 w.r.t \mathbf{y}_1 . Now consider program 1 with $A_{U,V}$ as measurement matrix, \mathbf{y}_1 as measurements, $\mathbf{x}_1 = -\Sigma_{X_0}$ as unknown vector and $\mathbf{w} = \Sigma_W$ as the

perturbation. Notice that $\mathbf{x}_1 + \mathbf{w}$ is as good as \mathbf{x}_1 w.r.t \mathbf{y}_1 . Also since \mathcal{A} has \mathbf{P}^e , $A_{U,V}$ has \mathbf{P} and thus V_1 holds for $A = A_{U,V}$. Using $V_1 \implies V_2$ we conclude

$$\|W\|_m = \|\mathbf{w}\|_v \leq h(\bar{\mathbf{x}}_1, \epsilon) = h(\Sigma(X_0), \epsilon) \quad (17)$$

It thus holds that $M_1 \implies M_2$.

Using similar arguments, we now show that $(V_1 \implies V_3) \implies (M_1 \implies M_3)$.

Assume $V_1 \implies V_3$ and that M_1 holds. Consider any $W \in \mathcal{N}(\mathcal{A})$ with SVD of $W = U\Sigma_W V^T$. Then, $\mathcal{A}(W) = A_{U,V}\Sigma(W) = 0$. Also $A_{U,V}$ satisfies \mathbf{P} and hence V_1 holds for $A = A_{U,V}$. Using $V_1 \implies V_3$, we find $\Sigma(W)$ satisfies \mathbf{Q} . Hence $M_1 \implies M_3$. ■

As can be seen from the Main Theorem, throughout the paper, we are dealing with a *strong* notion of recovery. By strong we mean \mathbf{P} guarantees recovery for all \mathbf{x}_0 (X_0) with sparsity (rank) at most k instead of for just a particular \mathbf{x}_0 (X_0). For example, results for the matrix completion problem in the literature don't have a *strong* recovery guarantee. On the other hand it is known that (good) RIP or SSP conditions guarantee recoverability for all vectors and yield *strong* recovery results.

In the next section, we show that the main theorem can be applied to RIP and SSP based recovery and thus the corresponding recovery results for matrix recovery easily follow.

A. Application of Main Theorem to RIP based recovery

We say that $f(\delta_{i_1}, \dots, \delta_{i_m}, \theta_{j_1, j'_1}, \dots, \theta_{j_n, j'_n}) \leq c$ is an RIP inequality where $c \geq 0$ is a constant and $f(\cdot)$ is an increasing function of its parameters (RIC and ROC) and $f(0, \dots, 0) = 0$. Let \mathcal{F} be a set of RIP inequalities namely f_1, \dots, f_N where k 'th inequality is of the form:

$$f_k(\delta_{i_{k,1}}, \dots, \delta_{i_{k,m_k}}, \theta_{j_{k,1}, j'_{k,1}}, \dots, \theta_{j_{k,n_k}, j'_{k,n_k}}) \leq c_k. \quad (18)$$

Lemma 5. *If $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ satisfies a set of (matrix) RIP and ROC inequalities \mathcal{F} , then for all unitary pairs (U, V) , $A_{U,V}$ will satisfy the same inequalities.*

Proof: Let $\delta'_k, \theta'_{k,k'}$ denote RIC and ROC of $A_{U,V}$ and $\delta_k, \theta_{k,k'}$ denote RIC and ROC of $\mathcal{A}(\cdot)$. Then we claim: $\delta'_k \leq \delta_k$ and $\theta'_{k,k'} \leq \theta_{k,k'}$. For any \mathbf{x} of $\|\mathbf{x}\|_0 \leq k$, let $X = U\text{diag}(\mathbf{x})V^T$. Using $\|\mathbf{x}\|_2 = \|X\|_F$ and $A_{U,V}\mathbf{x} = \mathcal{A}(X)$ we have:

$$(1 - \delta_k)\|\mathbf{x}\|_2^2 < \|\mathcal{A}(X)\|_F^2 = \|A_{U,V}\mathbf{x}\|_2^2 < (1 + \delta_k)\|\mathbf{x}\|_2^2$$

Hence $\delta'_k \leq \delta_k$. Similarly let \mathbf{x}, \mathbf{x}' have disjoint supports with sparsity at most k, k' respectively. Then obviously $X = U\text{diag}(\mathbf{x})V^T$ and $X' = U\text{diag}(\mathbf{x}')V^T$ satisfies the condition in ROC definition. Hence:

$$\begin{aligned} |\langle A_{U,V}\mathbf{x}, A_{U,V}\mathbf{x}' \rangle| &= |\langle \mathcal{A}(X), \mathcal{A}(X') \rangle| \\ &\leq \theta_{k,k'} \|X\|_F \|X'\|_F = \theta_{k,k'} \|\mathbf{x}\|_2 \|\mathbf{x}'\|_2 \end{aligned}$$

Hence $\theta'_{k,k'} \leq \theta_{k,k'}$. Thus, $A_{U,V}$ satisfies the set of inequalities \mathcal{F} as $f_i(\cdot)$'s are increasing function of δ_k 's and $\theta_{k,k'}$'s. ■

We can thus combine Lemma 5 and the main theorem to smoothly translate any implication of RIP for vector recovery to corresponding implication for matrix recovery. In particular, some typical RIP implications are as follows.

RIP implications for k -sparse recovery. ([3]) Suppose $\mathcal{A} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ satisfies, a set of RIP inequalities \mathcal{F} . Then for all \mathbf{x}_0 , $\|\mathbf{z}\|_2 \leq \epsilon$ and \mathbf{x}^* as good as \mathbf{x}_0 we have the following ℓ_2 and ℓ_1 robustness results,

$$\|\mathbf{x}_0 - \mathbf{x}^*\|_2 \leq \frac{C_1}{\sqrt{k}} \|\mathbf{x}_0 - \mathbf{x}_0^k\|_1 + C_2 \epsilon \quad (19)$$

$$\|\mathbf{x}_0 - \mathbf{x}^*\|_1 \leq C_3 \|\mathbf{x}_0 - \mathbf{x}_0^k\|_1 \quad (20)$$

for some constants $C_1, C_2, C_3 > 0$.

Now, using Lemma 5 and Theorem 1, we have the following implications for matrix recovery.

Lemma 6. *Suppose $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ satisfies the same inequalities \mathcal{F} as in (20). Then for all X_0 , $\|X\|_2 \leq \epsilon$ and X^* as good as X_0 we have the following Frobenius norm and nuclear norm robustness results,*

$$\|X_0 - X^*\|_F \leq \frac{C_1}{\sqrt{k}} \|X_0 - X_0^k\|_* + C_2 \epsilon \quad (21)$$

$$\|X_0 - X^*\|_* \leq C_3 \|X_0 - X_0^k\|_* \quad (22)$$

B. Application of Main theorem to SSP based recovery

Lemma 7. *Let $\Delta > 0$. If $\mathcal{A} : \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}^m$ satisfies Δ -SSP, then for all unitary pairs (U, V) , $A_{U,V}$ satisfies Δ -SSP.*

Proof: Consider any $\mathbf{w} \in \mathcal{N}(\mathcal{A}_{U,V})$. Then, $\mathcal{A}(U\text{diag}(\mathbf{w})V^T) = 0$ and therefore $W = U\text{diag}(\mathbf{w})V^T \in \mathcal{N}(\mathcal{A})$. Since \mathcal{A} satisfies Δ -SSP, we have $\frac{\|W\|_*}{\|W\|_F} = \frac{\|\mathbf{w}\|_1}{\|\mathbf{w}\|_2} \geq \sqrt{\Delta}$. Thus $A_{U,V}$ satisfies Δ -SSP. ■

Now we give the following SSP based result for matrices as an application of main theorem.

Theorem 2. *Consider program 2 with $\mathbf{z} = 0$, $\mathbf{y} = \mathcal{A}(X_0)$. Let X^* be as good as X_0 . Then if \mathcal{A} satisfies Δ -SSP with $\Delta > 4k$, it holds that*

$$\|X^* - X_0\|_* \leq C \|X_0 - X_0^k\|_* \quad (23)$$

where $C = \frac{2}{1 - 2\sqrt{k/\Delta}}$.

Note that the use of main theorem and Key Lemma simplifies the recovery analysis in [12] and also improves the sufficient condition of $k < \frac{\Delta}{6}$ in [12] to $k < \frac{\Delta}{4}$. This improved sufficient condition matches the sufficient condition given in [30] for the sparse vector recovery problem.

V. SIMPLIFIED ROBUSTNESS CONDITIONS

We show that various robustness conditions are equivalent to simple conditions on the measurement operator. The case of noiseless and perfectly sparse signals, is already given in Lemmas 3 and 4. Such simple conditions might be useful for analysis of nuclear norm minimization in later works. We state the conditions for matrices only; however, vector and matrix conditions will be identical (similar to Lemmas 3, 4) as one

can expect from Theorem 1. The proofs follow from simple algebraic manipulations with the help of Lemma 1.

Lemma 8. (Nuclear Norm Robustness for Matrices)

Assume $\epsilon = 0$ (no noise). Let $C > 1$ be constant. Then for any X_0 and any X^* as good as X_0 we are guaranteed to have:

$$\|X_0 - X^*\|_* < 2C\|X_0 - X_0^k\|_* \quad (24)$$

if and only if for all $W \in \mathcal{N}(\mathcal{A})$ we have:

$$\|W^k\|_* < \frac{C-1}{C+1}\|W - W^k\|_* \quad (25)$$

Lemma 9. (Frobenius Norm Robustness for Matrices)

Let $\epsilon = 0$. Then for any X_0 and X^* as good as X_0 ,

$$\|X_0 - X^*\|_F < \frac{C}{\sqrt{k}}\|X_0 - X_0^k\|_*, \quad (26)$$

if and only if for all $W \in \mathcal{N}(\mathcal{A})$,

$$\|W - W^k\|_* - \|W^k\|_* > \frac{2\sqrt{k}}{C}\|W\|_F \quad (27)$$

Lemma 10. (Matrix Noise Robustness) For any X_0 with $r(X_0) \leq k$, any $\|z\|_2 \leq \epsilon$ and any X^* as good as X_0 ,

$$\|X_0 - X^*\|_F < C\epsilon, \quad (28)$$

if and only if for any W with $\|W^k\|_* \geq \|W - W^k\|_*$,

$$\|W\|_F < \frac{C}{2}\|\mathcal{A}(W)\|_2 \quad (29)$$

VI. NULL SPACE BASED RECOVERY RESULT FOR SCHATTEN- p QUASI-NORM MINIMIZATION

In the previous sections, we stated the main theorem and considered its applications on RIP and SSP based conditions to show that results for recovery of sparse vectors can be analogously extended to recovery of low-rank matrices without making the recovery conditions stronger. In this section, we consider extending results from vectors to matrices using an algorithm different from ℓ_1 minimization or nuclear norm minimization.

The ℓ_p quasi-norm (with $0 < p < 1$) is given by $\|x\|_p^p = \sum_{i=1}^n |x_i|^p$. Note that for $p = 0$, this is nothing but the cardinality function. Thus it is natural to consider the minimization of the ℓ_p quasi-norm (as a surrogate for minimizing the cardinality function). Indeed, ℓ_p minimization has been a starting point for algorithms including Iterative Reweighted Least Squares [10] and Iterative Reweighted ℓ_1 minimization [15], [6]. Note that although ℓ_1 minimization is convex, ℓ_p minimization with $0 < p < 1$ is *non-convex*. However empirically, ℓ_p minimization based algorithms with $0 < p < 1$ have a better recovery performance as compared to ℓ_1 minimization (see e.g. [8],[15]). The recovery analysis of these algorithms has mostly been based on RIP. However Null space based recovery conditions analogous to those for ℓ_1 minimization have been given for ℓ_p minimization (see e.g. [17], [29]).

Let $\text{Tr}|A|^p = \text{Tr}(A^T A)^{\frac{p}{2}} = \sum_{i=1}^n \sigma_i^p(A)$ denote the Schatten- p quasi norm with $0 < p < 1$. Analogous to the

vector case, one can consider the minimization of the Schatten- p quasi-norm for the recovery of low-rank matrices,

$$\begin{aligned} & \text{minimize} \quad \text{Tr}|X|^p \\ & \text{subject to} \quad \mathcal{A}(X) = y \end{aligned} \quad (30)$$

where $y = \mathcal{A}(X_0)$ with X_0 being the low-rank solution we wish to recover. IRLS- p has been proposed as an algorithm to find a local minimum to (30) in [22]. However no null-space based recovery condition has been given for the recovery analysis of Schatten- p quasi norm minimization. We give such a condition below, after mentioning a few useful inequalities.

Lemma 11 ([28]). For any two matrices $A, B \in \mathbb{R}^{m \times n}$ it holds that $\sum_{i=1}^k (\sigma_i^p(A) - \sigma_i^p(B)) \leq \sum_{i=1}^k \sigma_i^p(A - B)$ for all $k = 1, 2, \dots, n$.

Note that the p quasi-norm of a vector satisfies the triangle inequality ($x, y \in \mathbb{R}^n$, $\sum_{i=1}^n |x_i + y_i|^p \leq \sum_{i=1}^n |x_i|^p + \sum_{i=1}^n |y_i|^p$). Lemma 11 generalizes this result to matrices.

Lemma 12 ([19]). For any two matrices A, B it holds that

$$\sigma_{t+s-1}(A + B) \leq \sigma_t(A) + \sigma_s(B) \quad (31)$$

where $t + s - 1 \leq n, t, s \geq 0$.

The following lemma easily follows as a consequence.

Lemma 13. For any two matrices, A, B with B of rank k and any $p > 0$,

$$\sum_{i=k+1}^{n-k} \sigma_i^p(A - B) \geq \sum_{i=2k+1}^n \sigma_i^p(A).$$

Theorem 3. Let $\text{rank}(X_0) = k$ and let \bar{X} denote the global minimizer of (30). A sufficient condition for $\bar{X} = X_0$ is that $\sum_{i=1}^{2k} \sigma_i^p(W) \leq \sum_{i=2k+1}^n \sigma_i^p(W)$ for all $W \in \mathcal{N}(\mathcal{A})$. A necessary condition for $\bar{X} = X_0$ is that $\sum_{i=1}^k \sigma_i^p(W) \leq \sum_{i=k+1}^n \sigma_i^p(W)$ for all $W \in \mathcal{N}(\mathcal{A})$.

Proof: (\Rightarrow) For any $W \in \mathcal{N}(\mathcal{A}) \setminus \{0\}$,

$$\begin{aligned} \text{Tr}|X_0 + W|^p &= \sum_{i=1}^k \sigma_i^p(X_0 + W) + \sum_{i=k+1}^n \sigma_i^p(X_0 + W) \\ &\geq \sum_{i=1}^k (\sigma_i^p(X_0) - \sigma_i^p(W)) + \sum_{i=2k+1}^n \sigma_i^p(W) \\ &\geq \text{Tr}|X_0|^p \end{aligned} \quad (32)$$

where the first inequality follows from Lemma 11 and Lemma 13. The necessary condition is easy to show analogous to the results for nuclear norm minimization. ■

Note that there is a gap between the necessary and sufficient conditions. We observe through numerical experiments that a better inequality such as $\sum_{i=1}^n \sigma_i^p(A - B) \geq \sum_{i=1}^n |\sigma_i^p(A) - \sigma_i^p(B)|$ seems to hold for any two matrices A, B . Note that this is in particular true for $p = 1$ (Lemma 1) and $p = 0$. If this inequality is proven true for all $0 < p < 1$, then we could

bridge the gap between necessity and sufficiency in Theorem 3. Thus we have that singular value inequalities including those in Lemma 1 and Lemma 13, 11 play a fundamental role in extending recovery results from vectors to matrices. Although, our condition is not tight, we can still use the Theorems 1 and 3 to conclude the following:

Lemma 14. Assume property **S** on matrices $\mathbb{R}^n \rightarrow \mathbb{R}^m$ implies perfect recovery of all vectors with sparsity at most $2k$ via ℓ_p quasi-norm minimization where $0 < p < 1$. Then \mathbf{S}^e implies perfect recovery of all matrices with rank at most k via Schatten- p quasi-norm minimization.

Proof Idea: Since **S** implies perfect recovery of all vectors with sparsity at most $2k$, it necessarily implies a null space property (call it **Q**) similar to the one given in Lemma 3, (10) but with p in the exponent (see e.g. [29]) and k replaced by $2k$. Now the main theorem, (15), combined with Theorem 3 implies that \mathbf{S}^e is sufficient for perfect recovery of rank k matrices.

In particular, using Lemma 14, we can conclude, any set of RIP conditions that are sufficient for recovery of vectors of sparsity up to $2k$ via ℓ_p minimization, are also sufficient for recovery of matrices of rank up to k via Schatten- p minimization. As an immediate consequence it follows that the results in [7], [15] can be easily extended.

VII. CONCLUSIONS

We presented a general result stating that the extension of any sufficient condition for the recovery of sparse vectors using ℓ_1 minimization is also sufficient for the recovery of low-rank matrices using nuclear norm minimization. Consequently, we have that the best known RIP-based recovery conditions of $\delta_k < 0.307$ (and $\delta_{2k} < 0.472$) for sparse vector recovery are also sufficient for low-rank matrix recovery. Note that our result shows there is no “gap” between the recovery conditions for vectors and matrices, and we do not lose a factor of 2 in the rank of matrices that can be recovered, as might be suggested by existing analysis (e.g. [27]).

We showed that a Null-space based sufficient condition (Spherical Section Property) given in [30] easily extends to the matrix case, tightening the existing conditions for low-rank matrix recovery [12]. Finally, we gave null-space based conditions for recovery using Schatten- p quasi-norm minimization and showed that RIP based conditions for ℓ_p minimization extend to the matrix case. We note that all of these results rely on the ability to “vectorize” matrices through the use of key singular value inequalities including Lemma 1, 11.

REFERENCES

- [1] T.T. Cai, G. Xu, and J. Zhang. New bounds for restricted isometry constants. *IEEE Tran. Info. Theory*, 56:4388–4394, 2010.
- [2] T.T. Cai, G. Xu, and J. Zhang. Shifting inequality and recovery of sparse signals. *IEEE Tran. Signal Processing*, 58:1300–1308, 2010.
- [3] E.J. Candes. The restricted isometry property and its implications for compressed sensing. *Academie des Sciences*, 346:589–592, 2008.
- [4] E.J. Candes and Y. Plan. Tight oracle bounds for low-rank matrix recovery from a minimal number of random measurements. 2009.
- [5] E.J. Candes and T. Tao. Decoding by linear programming. *IEEE Trans Info. Theory*, 51:4203–4215, 2004.

- [6] E.J. Candes, M.B. Wakin, and S. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14:877–905, 2008.
- [7] R. Chartrand and V. Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24:1–14, 2008.
- [8] R. Chartrand and W. Yin. Iteratively reweighted algorithms for compressive sensing. In *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2008.
- [9] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *Journal of the American Mathematical Society*, 22(1):211–231, 2009.
- [10] I. Daubechies, R. DeVore, M. Fornasier, and C.S. Gunturk. Iteratively re-weighted least squares minimization for sparse recovery. *Commun. Pure Appl. Math*, 63(1), 2010.
- [11] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. on Info. Theory*, 47(7), 2001.
- [12] K. Dvijotham and M. Fazel. A nullspace analysis of the nuclear norm heuristic for rank minimization. In *Proc. of ICASSP*, 2010.
- [13] M. Elad and A. M. Bruckstein. A generalized uncertainty principle and sparse representation in pairs of bases. *IEEE Trans. on Info. Theory*, 48(8), 2002.
- [14] A. Feuer and A. Nemirovsky. On sparse representation in pairs of bases. *IEEE Trans. on Info. Theory*, 49(6), 2003.
- [15] S. Foucart and M.J. Lai. Sparsest solutions of underdetermined linear systems via ℓ_q minimization for $0 < q < 1$. *Applied and Computational Harmonic Analysis*, 26/3:395–407, 2009.
- [16] R. Garg and R.Khandekar. Gradient descent with sparsification: An iterative algorithm for sparse recovery with restricted isometry property. In *Proc. of 26th Annual ICML*, 2009.
- [17] R. Gribonval and M. Nielsen. Sparse representations in unions of bases. *IEEE Trans. on Info. Theory*, 49(12):3320–3325, 2003.
- [18] R.A. Horn and C.R. Johnson. *Matrix Analysis*. 1990.
- [19] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis*. 1990.
- [20] K. Lee and Y. Bresler. Admira: Atomic decomposition for minimum rank approximation. 2009. Available at <http://arxiv.org/abs/0905.0044>.
- [21] R. Meka, P. Jain, and I.S. Dhillon. Guaranteed rank minimization via singular value projection. In *Proc. of NIPS*, 2010.
- [22] K. Mohan and M. Fazel. Iterative reweighted least squares for matrix rank minimization. In *Proc. Allerton conference on Controls and communications*, 2010.
- [23] K. Mohan and M. Fazel. New restricted isometry results for noisy recovery. In *Proc. International Symposium Information Theory*, 2010.
- [24] D. Needell. Noisy signal recovery via iterative reweighted ℓ_1 -minimization. In *Proc. Asilomar conference on Signals, Systems and Computers*, 2009. Available at <http://arxiv.org/abs/0904.3780>.
- [25] D. Needell and J.A. Tropp. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Applied Computational Harmonic Analysis*, 26:301 – 321, 2008.
- [26] S. Oymak and B. Hassibi. New null space results and recovery thresholds for matrix rank minimization. 2010. Available at <http://arxiv.org/abs/1011.6326>.
- [27] B. Recht, M. Fazel, and P. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52:471–501, 2010.
- [28] M. Uchiyama. Subadditivity of eigenvalue sums. *Proc. American Mathematical Society*, pages 1405–1412, 2005.
- [29] M. Wang, W. Xu, and A. Tang. On the performance of sparse recovery via ℓ_p minimization. 2010. Available at <http://arxiv.org/abs/1011.5936>.
- [30] Y. Zhang. Theory of compressive sensing via ℓ_1 minimization: A Non-RIP analysis and extensions. 2008. Technical Report TR08-11 revised. Available at http://www.caam.rice.edu/~zhang/reports/tr0811_revised.pdf.